DOCUMENT RESUME

ED 226 088 UD 022 656

AUTHOR Hall, Janie L.
TITLE A Study of Mental Ability Testing and Its Implications for the Oklahoma City Public Schools.
INSTITUTION Oklahoma City Public Schools, OK. Dept. of Planning, Research, and Evaluation.
PUB DATE Mar 81
NOTE 37p.
PUB TYPE Reports - General (140)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cognitive Ability; *Cost Effectiveness; Elementary Secondary Education; *Intelligence Tests; Minority Groups; Relevance (Education); *Test Bias; Test Reliability; *Test Validity; Urban Schools
IDENTIFIERS *Oklahoma City Public Schools; *Otis Lennon Mental Ability Test

ABSTRACT

This study follows a 1980 moratorium on group mental ability testing called by the district's superintendent when questions relating to the informational value and cost-effectiveness of the Otis Lennon Mental Ability Test (OLMA) were raised by the Oklahoma City Public School District. Criticisms of intelligence tests and relevant issues are discussed in the following areas: nature of characteristics measured, cultural bias, extent of use, purposes which tests serve, misuses, psychometric concerns, cost, legislation affecting mental ability testing, and committee input. This report recommends discontinuing CLMA tests; ordering no new aptitude tests; and relying on other sources of information such as criterion-referenced tests, achievement tests, individual intelligence tests, and the "time-to-learn" concepts. (Author/WAM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED226088

A STUDY OF MENTAL ABILITY TESTING

AND ITS IMPLICATIONS FOR THE OKALHOMA CITY PUBLIC SCHOOLS

Janie L. Hall

Standardized Test Specialist

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

✓ This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Janie L. Hall
Oklahoma City Public Schools

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

March, 1981
Department of Planning, Research, and Evaluation
Pat Watson, Director
Oklahoma City Public Schools
Oklahoma City, Oklahoma

UD 020 956

Foreward

The Oklahoma City Public School District administered the Otis Lennon Mental Ability Test to students in grades 3, 5, 7, 9, and 11 from 1967 through 1979. The test was used primarily for grouping students and for estimating potential. However, in recent years, concerns were raised related to the appropriateness of these uses. The value of the information gained was questionable when compared with the time and monies expended. Additionally, some theorists question the concept of mental ability testing.

Therefore the Superintendent's staff called a moratorium on mental ability testing in the fall of 1980 until a study of the matter could be conducted. The following report and the summary which precedes it are the result of that investigation.

SUMMARY

A STUDY OF MENTAL ABILITY TESTING AND ITS IMPLICATIONS FOR THE OKLAHOMA CITY SCHOOLS

By
Janie Hall
Standardized Test Specialist

Purpose of the Study

The purpose of the study is to review theory and practice in group mental ability testing and relate these to the needs of the Oklahoma City School District. It is expected that the information obtained will aid in determining whether district-wide mental ability testing should continue and, if so, to assist in test selection. Criticisms of tests and issues which seem relevant to our situation are grouped under the following headings: nature of characteristics measured, cultural bias, extent of use, purposes which tests serve, misuses, psychometric concerns, cost, legislation affecting ability testing, and committee input.

Nature of Characteristics Measured

While there continues to be a lack of agreement among theorists about what is measured by intelligence tests, there is growing agreement that whatever is measured is responsive to training. Other issues which are being given attention include emphasis on measuring the separate components of intelligence, and on discontinuing use of terms like "intelligence tests" and "IQ tests". Several of the major test producers are responding with new ability tests.

Cultural Bias

There is little agreement among researchers about whether tests are excessively reflective of middle class values. One approach states that if the purpose of such tests is to aid in moving individuals into a given culture then the test should reflect the values of that culture. While a number of courts have ruled tests to be inappropriate for minorities, some researchers believe that judgements can only be made about specific tests with specific groups. Some maintain that culture-reduced tests discriminate against those they are intended to help by making them ineligible for special education. A review of U.S. census figures indicates that norming samples which are selected according to the ethnic composition of the general population may not be representative of the Oklahoma City schools.

Extent of Use

Figures regarding extent of use across the country vary from study to study. In one, the use of ability tests by large districts decreased from 41% at some grade levels to 77% at others. One study found 43% of districts choosing to use ability tests and an equal percent choosing not to, while another study found 43% of respondents required administering ability tests and 35% were not using them; the remaining 22% made them optional. Two school districts found school personnel expressing a need for the tests, but failed to find corresponding evidence of use. One city which discontinued use has returned to limited use. There are a number of districts which are strongly for and against ability tests, but responses to surveys seem to reflect considerable confusion on the part of others.

Purposes which Intelligence Tests Serve

Purposes most often stated by large urban districts, (including Oklahoma City) are: for placement, eligibility, grouping, estimate of potential, comparing achievement with potential, for parent conferences for lesson planning, for recommending for further testing. One theorist states that group tests should only be used for describing school populations (several school districts expressed similar views), and that the value of information gained by mass testing cannot justify the cost.

Misuses

A number of the uses pointed out by school districts are considered misuses by theorists. Two of these are ability grouping and estimation of potential. Other misuses cited are overreliance on scores, labelling, nonstandardized testing conditions and lack of consolidation. Possible replacements for mental ability tests which are suggested include achievement tests, criterion-referenced tests, and time to learn measures.

Psychometric Concerns

Norms for the test currently used in Oklahoma City are 23 years old and teachers have noticed problems with outmoded use of words. However, the effect of this on district scores is not directly observable; a more serious problem may be the difference between the ethnic composition of Oklahoma City School District and the national population. Regarding stability over time, group scores appear quite stable, but a high percentage of individual students vary by 10 points or more over two-year intervals. This seems to bear out comments by theorists that group tests should be used only for describing district populations. Correlations with achievement test results are high; however, standard errors of estimate are large suggesting the possibility of wide discrepancies for individuals.

Cost

Cost of administering the 1967 OLMA runs about $4501.00 per year; approximately 30,714 hours are spent each year in the testing process by students and personnel. If new tests are purchased, initial cost to test grades 1-11 will be approximately $24,630 with a maintenance cost of about $11,778 and involving @ 16,00 to @ 18,000 hours.

Legislation Affecting Testing in OKC

A bill was proposed in the Oklahoma Legislature during 1981 which would require a district testing program, designate the tests to be used, outline regulations for administration, and require an annual report. However, the proposal has been defeated for this session.

Input from Testing Committees

After reviewing information gained in the study and receiving input from school testing committees, the following recommendations are made by the Steering Committee:

1. The OKC district should stop administering the Otis Lennon Mental Ability Test because the norms are 23 years old and because they are not validated with the CAT norms.

2. No new aptitude test should be purchased because of the high correlation between CAT and the companion aptitude test; i.e., the two tests measure the same thing.

Recommendations

It is recommended that:

- The OLMA not be used
- No new test be ordered at this time
- That teachers be given assistance in using other sources of information.

A Study of Mental Ability Testing and Its Implications for the Oklahoma City Schools

The Otis-Lennon Mental Ability Test (OLMA), published by the Psychological Corporation in 1967, was selected by a testing committee in 1973, and continued to be used in certain grades in the Oklahoma City Public Schools through the fall of 1979. Once during that period (1975) the testing committee reviewed a number of other tests because it was felt that the OLMA might contain cultural bias and because members preferred a test which provided verbal and quantitative scores. Failing to find a more satisfactory test, use of the OLMA was continued. In 1980, a moratorium on use of the OLMA was called by the superintendent and his staff until such time as a study of current theory, practices and needs could be completed.

This study is designed to explore those areas and to aid in answering the following questions:

1. Should a group mental ability test be administered district-wide; and,
2. If so, to what grades, and using what test?

Criticisms and issues which seem relevant to our situation are grouped under the following headings for discussion: nature of the characteristics measured, cultural bias, extent of use, purposes which tests serve, psychometric concerns, cost, and legislation effecting ability testing, and input from testing committees.

Nature of Characteristics Measured

Trends in Test Theory

Discussions of mental ability tests (also referred to as intelligence or IQ tests) frequently do not differentiate between group and individual tests. The two types rest on the same theoretical foundations, although the superiority of individual tests is generally acknowledged (Jensen, 1980; Vandiver and Vandiver, 1979). Because this paper is primarily concerned with group tests, they will be identified in the discussion whenever possible.

Intelligence tests were originally designed to sample a variety of functions believed to estimate an individual's general intellectual level and which Charles Spearman (1904) defined as general factor of "g". This general factor is said to be measured by all intercorrelated tests (Jensen, 1980). However, Blum (1978) views the entire hypothesis of general factor as unprovable and recommends the abandonment of mental ability tests. Wallbrown (1979) and Jensen (1980) are less harsh but suggest that we do not know what we mean by general factor.

Some group tests (Otis and Lennon, 1969) are said to measure this broadly based general intellectual ability and to reflect a combination of innate and learned abilities. Conversely, some writers insist that intelligence (IQ) tests do not measure innate potential (Reschley, 1981). There is growing agreement that whatever is measured by intelligence tests is responsive to training. (Anastasi, 1980; Fine, 1975; Bloom, 1976; Jacobs and Vandiver, 1971; Olton and Crutchfield, 1969; Resnick and Glaser 1976).

Much of current and predicted theoretical interest emphasizes the separate components of intelligence (Horn, 1979; Thorndike, 1971; Anastasi, 1980; Gray, 1978). Some maintain that different components are valued by different cultures (Horn, 1979; Sternberg, 1979; Anastasi, 1980). In the case of intelligence testing in schools, this is the academic component as opposed to the real life component (Neisser, 1979). Sternberg (1979) agrees that different cultures may value different components but argues that all people seem to have the same components.

And so the discussion of what is being measured goes on. Berger (1973) and Olson (1976) say that it is the extent to which an individual has internalized the cognitive requirements of modern technological-rationalistic world. Others (Weber, 1979) argue the relationship is high between verbal skills and measured intelligence, particularly when group tests are used. Weber adds that what is measured is a kind of general academic achievement in standard English. He believes the implicit assumption is that all persons being tested have had the same opportunity to learn the kinds of things included on the test, and that such is not the case unless tasks are entirely new to all subjects.

Perhaps modern views relative to what IQ tests measure can be summarized by quoting a statement which Dr. Kathy Osborn (1981) proposed should be incorporated in the write-up of every student tested for possible placement in special education. Dr. Osborn, basing her views on the remarks of Reschley (1981), had individual tests in mind. But the cautions are even

more appropriate for group tests. She suggests it be noted, that

> "IQ tests measure only a portion of the competencies involved in human intelligence. The IQ results are best seen as predicting performance in school and reflecting the degree to which children have mastered predominant cultural symbols and values. This is useful information, but it is also limited. Further cautions - IQ tests do not measure innate genetic capacity and the scores are not fixed. Some persons do exhibit significant increases or decreases in their measured IQ."

Responses by Group Test Developers

Test developers, as one might expect, are reacting to changing views. William Turnbull, president of Educational Testing Service (1979), indicated that in the future tests are more likely to measure more aspects of ability, the tasks will be less abstract and closer to life experiences, and that each measured ability will yield a separate score. He expects that norms will be related to cultural and linguistic background.

Increased interest in reporting on the components of intellect is reflected in the Developing Cognitive Abilities Test (Wick, et. al., 1980) It is said to provide information on some levels of Bloom's taxonomy (Bloom, et. al., 1956).

The new Otis-Lennon Scholastic Ability Test, OLSAT (Otis and Lennon, 1979) reflects a revival of the term scholastic aptitude in recognition of the fact that what is being measured consists of the components related to success in formal education. Some hope that the term "scholastic aptitude" will not have the negative connotations that the term "intelligence" does.

But Lennon (1978) is skeptical and wonders if older terms can be successfully avoided and whether new ones will be more accurately used than the old ones.

Perhaps CTB/McGraw Hill has been the most innovative. They advertise that their new Short Form Test of Academic Aptitude (technical manual not yet available) measures a minimum of achievement. For example, in the memory section, real but little-known works are presented. This seems to be an attempt to address views, such as Weber's (1979), that material should be equally new to all subjects.

Cultural Bias

It has been argued by some that intelligence tests (group and individual) are excessively reflective of middle class values, that content is inappropriate for minorities, and that minorities are underrepresented in norming samples. These points are addressed separately as follows.

Excessively Reflect Middle Class Values

Empirical studies by Coffman (1967), Deutsch, et. al. (1964), and Jones, et. al. (1971) using group mental ability tests seem to show that the tests are as valid for predictive purposes for individuals in one social class as another while others have demonstrated conflicting results (Steffensen, et. al.).

Lennon (1978) says that there need be no concern for cultural sources of variance as long as no inferences are made about potential or inherited capacity. Interestingly, the Manual for Administration of OLMA (1967,

p.18) cites "Evaluation of learning potential..." as a recommended use for test results while the manual for the 1979 edition (p. 20) presents a slightly different recommendation, "... estimating the pupil's learning ability or current readiness to deal with academic work for his or her particular grade or class."

Views expressed by Anastasi (1980), Weber (1978), and Lennon (1978) seem to be a practical response to the issue. They suggest that values reflected by a test should depend on the purpose of the test. If that purpose is to describe a given culture then it must reflect the values of that culture, but if the purpose is to move an individual into another culture, then tasks must reflect the values of the receiving culture.

Content Inappropriate for Minorities

In 1967, Circuit Judge J.S. Wright of Washington, DC, stated that test questions which were inappropriate to the background of disadvantaged children and comparisons to a national norming group make it likely that a majority of District of Columbia children are in danger of being wrongly labelled.

A much publicized case in California, Larry P., et. al. v. Wilson Riles, involved the accusation that an individual test (WISC-R) is biased against the culture and experience of black children. The result was a moratorium on the administration of IQ tests of all children in the state.

Prior to the California ruling, a number of cities, including New York, Chicago, Los Angeles, and Houston had already curtailed or eliminated

the use of standardized tests of academic aptitude and achievement and were seeking other means of assessment to replace them, usually through use of criterion-referenced tests (Jensen, 1980). Houston has since readopted a limited amount of mental ability testing.

Clarizo (1979) and Jensen (1980), after citing numerous studies of possible dialect bias against lower class black children, concluded that there is little, if any, empirical support for the accusation. And they maintain the evidence is too skimpy to permit generalizations regarding minority groups in general. Clarizo cautions that linguistic bias may be a function of specific tests and specific groups whose native tongue is not English, and that each test must be evaluated in its own right for use with minority groups. Jacobs (1980) adds that highly mobile students and returning dropouts are also groups needing special attention and that criterion-referenced tests are the tests of choice. Horn (1979) maintains there can be culture-free tests of components but not perfectly culture-free tests of intelligence. Lennon's rebuttal (1978) is to state that mental ability tests are accurately descriptive of the here and now on tasks important to learning.

Norms are Underrepresentative of Minorities

Flaugher (1978) and Clarizo (1979) defend established testing practices against more recently developed culrure-reduced testing systems such as Mercer's System of Multicultural Pluralistic Assessment (1979). It is Clarizo's belief that the scales which use within group norms discriminate

against the students they are intended to help by making those students ineligible for special services. He maintains that since a child from a culturally different background must function within the predominant culture, it seems reasonable to know how the disadvantaged children function in relation to majority children and that current IQ tests do what they are proposed to do.

The Psychological Corporation and CTB/McGraw Hill report that the proportion of minorities in their recent norming samples are as representative of the proportions they occupy in the population as possible, given the fact that participation in national standardizations is voluntary.

The 1967 OLMA technical manual includes no racial breakdown of the norming group, but Table I compares the 1979 OLSAT sample with the national population (1970 Census Bureau figures) and the OKC population (second quarter 1980-81).

Table 1

Ethnic Breakdown of Pupils in OLSAT Sample,
National Population and OKC Population

Ethnic Breakdown	White	Black	Hispanic	Other
OLSAT Sample	74%	20%	4%	2%
National Population	76%	16%	6%	2%
OKC Population	54%	37%	4%	6%

While the OLSAT norming sample is reasonably representative of the national population, it is clearly less representative of the Oklahoma City School District.

Extent of Use

Nationally

Lennon, in his 1978 address to the National Council on Measurement in Education, mentioned a poll of the membership which indicated that 9% predicted an increase in attention to intelligence testing in the next ten years, 42% predicted less attention, and 42% looked for no change.

A survey of large cities and counties conducted by Dimengo, in 1978, indicated that the use of group mental ability tests had declined in grades K-12 since 1964. The percent of decline is shown in Table 2. (The use of achievement tests had not changed, and 90.3% of respondents indicated that they used criterion-referenced tests in 1978).

Four years later, Hall (1981) surveyed 63 large urban districts; of the 49 respondents, 21 (42.9%) indicated that they used mental ability tests as a matter of choice. One other indicated that it was done in reponse to state mandate. Twenty-one districts (42.9%) chose not to use them and six California districts (12.2%) said state law forbade it. They did not indicate agreement or disagreement. The 1981 users are shown by the same levels as those reported by Dimengo in Table 2. Many of the same districts were in both groups.

Table 2

Results of Dimengo and Hall Surveys
Percent of Respondents Using Group Mental Ability Tests
Large City and County School Districts*

	Dimengo 1978 (N=77)			Hall, 1981 (N=49)
Grade Range	Percent of Respondents Using		Percent of Decline	Percent of Respondents Using
	1964	1978		1981
K-3	90.9%	32.4%	58.5%	34.7%
4-6	100.0	23.4	76.6	40.8
7-9	100.0	35.1	64.9	34.7
10-12	85.4	44.2	41.2	18.4

* 35 Districts were in both groups

One large urban district indicated (Hall 1981) that, when mental ability tests were offered, there were few requests, and no complaints when they were no longer offered, even though 82% of respondents to a survey expressed the belief that an intelligence test should be administered. Another district indicated that after mental ability testing was stopped there were complaints from school psychologists who felt that one aspect of student information was being missed.

In a survey of the National Education Association by Kovacs (1980) 805 member districts responded. Responses by systems with 25,000 students or more are reported here. Forty-three percent (43%) said mental ability tests were required (by state or system mandate), 22% made them optional, and 35% did not use them. Again, it is expected that some districts which responded to this study were also included in the previous two surveys.

Oklahoma City

The OLMA has been administered to grades 3, 5, 7, 9, and 11 since 1977. In 1980, Hall sent questionnaires to a sample of teachers, all principals, and all instructional administrators to learn the extent of use and the ways scores are used. The 214 responses received were not sufficient in number to be considered representative, but showed that 30.1% of respondents did not use the test results, while the remaining respondents (69.9%) said they used results for a variety of purposes which will be identified below in the "purposes" section. There was little agreement on the grades that should be tested.

A follow-up study by Hall in 1981 presented the questions asked by this study to all district staff members in instruction-related roles. The 1322 responses (of which 718 were teachers) were sufficient to be considered statistically representative. Of these, 66% indicated that a mental ability test should be given district-wide; 33% said "no"; and 1% had no opinion. There was still no agreement about which grades to test. The largest percent of agreement was 3.2% who said "all"; there were more than 100 other combinations of grades ranging in number from one to eight. It was learned that some respondents to this survey thought that the term "mental ability test" included achievement tests and responded accordingly. However, after talking with testing committees and noting comments which accompanied responses, the researcher feels that if incorrect responses were eliminated, the proportions of responses would remain essentially the same.

The individual responsible for group testing has received no complaints since mental ability testing was suspended. A small number of teachers have mentioned that they were using scores from previous years.

Purposes Which Intelligence Tests Serve

Uses of Intelligence Tests, in the Literature

Resnick (1979) separated uses into those which assist in management of instruction and those addressing public accountability. Under management, tests serve a sorting function (eligibility for programs), a grading function (interpreting achievement test scores), and a monitoring function (although there is no test which does this, according to Resnick). The

public accountability function is brought into use when schools use IQ scores to set limits on public expectations for achievement. This is usually the result of public pressure for higher achievement.

Jensen (1980) describes what he calls the one valid use of group mental ability tests as that of describing school populations. Uses cited by test publishers included prediction of achievement scores, identification for special programs, evaluating instructional programs, and grouping for instruction.

Uses of Group Tests by Other Districts

Other large urban districts indicated (Hall, 1981) that they used intelligence test scores as estimates of ability and potential, as broad screening instruments, as aids in planning, for comparison with achievement scores, as aids in counseling, as group assessment tools, and to determine eligibility for gifted/talented programs.

Uses of Group Tests in Oklahoma City

In the 1980 study by Hall, the following uses were indicated by 214 respondents:

- For placement, eligibility for special programs, and grouping in class (32.4%)
- Did not use (30.1%)
- To estimate potential (28.7%)

·To compare achievement with potential (31.7%)

·When considering students for further testing (15.4%)

·For parent conferences (13.2%)

·For lesson planning (5.1%)

·As part of overall picture of student (2.2%)

·(Others by 1% or less)

Comments in the 1981 study stated the same uses. In both studies there appeared to be overlaps in the responses and some individuals gave more than one response.

One use that needs to be specifically noted is that of determining eligibility for gifted and talented programs. Information was available for 781 participants (in 19 OKC schools) in 1979-80. Two hundred eighty-two (282) were in grades which were tested with the OLMA in the fall of 1979; of these, 267 students actually had scores. In 156 (56%) of the cases, the OLMA was cited as one of the criteria for placement. In one school (59 students), it was given as the major criterion. No schools used it as the only criterion. Achievement scores were part of the criteria in most cases and the only criterion in many cases. In a few cases, teacher and/or parent nomination alone was used.

Misuses

Misuses Noted in the Literature

Ability grouping. Fine (1978) maintains that the use of ability tests for grouping establishes a vicious cycle; programs match performance which in

turn matches program. Thorndike (1971) reports research related to ability grouping which suggests that gains shown by able students are accompanied by losses by less able students and an over-all loss on a school-wide basis. Jensen (1980) disagrees somewhat, stating that grouping has positive effects unless students are in the same groups for long periods.

Estimation of potential. This is no longer considered a valid use of mental ability tests (Jensen, 1980; Lennon, 1978; Anastasi, 1980). Rather, tests measure current performance which is an accumulation of experiences.

Over-reliance on scores. Jensen (1980) and Lloyd (1979) believe that too much emphasis and confidence is placed in scores.

Discrimination. Lloyd (1979) accuses tests of discriminating against low socioeconomic groups. Along with this he notes the failure of test users to realize the punitive impact of environment on test performance. Cottle, (1977) emphatically states that much of the intelligence testing that is done results in labelling which is damaging to student's self-esteem. To this Jensen (1980) responds that all tests discriminate, that objections to labelling involve the fact that tests often discriminate among certain groups, and that as long as individual differences exist, labelling will exist.

Nonstandardized conditions. Group tests are often administered under nonstandardized conditions requiring reading, and do not allow for trained observations (Vandiver and Vandiver, 1979).

Too many tests. Jacobs (1980) believes that researchers and administrators are faced with too many tests from which to choose and suggests consolidation.

Recommendations for Replacement Noted in the Literature

Jensen (1980, p. 740) believes that the information gained is too general to be useful in planning instruction and that what is gained is redundant to what is already known. He indicates, "... there is no routine use of mental ability tests in the schools for which well-designed achievement tests would not better serve the same purpose." He adds that the value of the information gained by mass testing cannot justify the cost.

Resnick (1979) states that all current tracking decisions could be made with achievement tests, that instructional monitoring is best done with criterion-referenced tests, and that IQ tests could be dropped with little effect on formal and informal grouping practices.

Glaser (1979) says tests reflect a restrictive assessment of abilities to learn and so have limited value to guidance and improvement of student achievement.

A study by Gettinger and White (1979) showed a higher correlation between time to learn and achievement than between IQ scores and achievement. This seems to suggest an area for investigation by schools using criterion-referenced testing.

Psychometric Concerns

Norms, 1967 OLMA

Norming dates. Of concern to this discussion is the fact that norming for the OLMA took place 23 years ago. Some Oklahoma City teachers have called attention to seldom-used words which students frequently misunderstand. For example: a vocabulary item on the third grade test shows a balloon, a laundry hamper, a punching bag, and a van. The student is asked to identify the "hamper". Teachers report that students frequently identify the van thinking that what is meant is a "camper".

Table 3 shows district-wide scores since 1973. The average differences across grades show more losses than gains, but not in a pattern which might suggest that distance from norming year is a factor. Rather, the average differences for each grade, across years, show increasing losses at the higher grade levels.

Table 3

National Percentile Rank
Oklahoma City School District, (1973-1979)

Grade	1973	1974	1975	1976	1977	1978	1979	Average Each Grade
3		48			48	50 (+2)*	48 (-2)*	1.33
5	44	43 (+7)*	41 (-2)*	42 (+1)*	40 (+1)*	38 (-2)	37 (-1)	.17
7	36	43 (+2)	41 (-2)	41 (0)	39	42 (-3)	41 (-2)	-.33
9	41				37	37 (0)	36 (-1)	-. 5
11	34				41	37 (-4)		-. 4
Average Diff. Across Grades		(+4.5)	(-2)	(+.5)	(-2.5)	(-.2)	(-1.6)	

*Difference when compared with previous year

Representativeness of Norms. An additional problem for Oklahoma City may be the difference in ethnic composition when compared with the national population (Table 1, p. 10). Any norming sample based on that population would seem less representative of this district.

No comments were found in the literature regarding use of ability tests and achievement tests which were normed using different samples. However, several school districts recommended comparing ability with achievement only where the two tests were normed on the same sample.

Temporal Stability of OLMA Scores

Jensen (1980) maintains that correlations are high over time (.83-.90), meaning that there is a strong relationship between student's scores on the same test taken at different times. He further believes that group means and rank order vary little from time to time. Others argue that intelligence scores tend to drop during adolescense (Kaufman, 1979), while others say that the cumulative nature of skill acquisition indicates that scores should increase

Table 4 provides a look at scores for a random sample of students (stratified by school) who took the OLMA in grades 5, 7, 9, and 11. Our students do not provide a response to the argument about the cumulative nature of skill acquisiton; the mean is smallest at grade five, becomes progressively larger through nine, but drops at eleven.

Values given for "r" show a strong relationship exists between tests at two-year intervals, but the relationship decreases slightly as time between

tests increases. Research done with the OLMA norming group over a six-year period (Robertson and Mitchell, 1975) showed similar findings with fifth grade and above. Correlations with younger students led them to surmise that age at the time of original testing is a more important variable than length of interval between tests. This is believed to be due to rapid development in the early years.

Table 4

Summary of Temporal Stability Data
Sample of Oklahoma City Students
OLMA scores, 1973-1979
N=345

Time Interval	Grades	First Mean	Second Mean	Difference in means	r	r^2	Difference ± 10
2 yrs.	5-7	39.61	41.92	2.31	.88*	77%	34%
2 yrs.	7-9	41.92	45.71	3.79	.88*	77%	40%
2 yrs.	9-11	45.72	43.10	-2.61	.86*	73%	42%
4 yrs.	5-9	39.61	45.71	6.10	.84*	71%	67%
4 yrs.	7-11	41.92	43.10	1.18	.83*	69%	41%
6 yrs.	5-11	39.61	43.10	3.49	.77*	59%	39%

* $p < .0001$
(standard deviations ranged from 26.84 to 27.22)

Values under "r^2" indicate the percent of variability in the second scores which is accounted for by the first scores. For example, 77% of the variability in seventh grade scores is accounted for by fifth grade scores and the remaining 23% is due to other factors. So we may say that we have a fairly good predictive ability over two-year intervals for groups. However, the "difference" column indicates a relatively large percent of individuals had scores deviating by 10 points or more from interval to interval. This is confirmed by the large standard deviations.

These differences seem important when one considers our use of mental ability scores, and the problem is compounded by testing only every other year. That is, a score for a given student may be quite unrelated to ability a year from the time it is obtained and yet still be used for estimating potential and for comparison with achievement scores. And contrary to Jensen's statement (1980), our rank order varies considerably from test to test; this may be important when ability scores are used for placement in gifted and talented programs.

Correlation between OLMA and Achievement Tests

It is generally acknowledged that intelligence tests (group and individual) are highly correlated with each other and with achievement tests (Jensen, 1980; Neisser, 1979; Lennon, 1978; Resnick, 1979). Under certain circumstances the same test items may appear on both mental ability and achievement tests (Thorndike, 1971), and some have said that intelligence tests are achievement tests with a predictive purpose (Horn, 1979).

The relationship between Oklahoma City's intelligence test scores and achievement test scores for a stratified random sample of students is shown in Table 6. The sample is from 1978-80 scores because same-year correlations could not be made between OLMA and achievement scores for the original sample.

Table 6

Correlations between OLMA scores and CAT scores
Oklahoma City Schools
Grades 3, 5, 11
1979-80

Grade	N	OLMA/Reading			OLMA/Math			Reading/Math		
		r	r^2	est.	r	r^2	est.	r	r^2	est.
3	368	.64*	.41%	8.85	.67*	.45%	8.51	.81	.66%	9.58
5	382	.85*	.72%	8.40	.81*	.66%	9.37	.78	.61%	10.29
11	350	.83*	.69%	8.77	.82*	.67%	8.93	.76	.58%	9.10

p < .001

The figures for "r" are the degree of correlation between the scores; the higher the number, the stronger the relationship. Group intelligence tests are believed to be highly related to reading skills; but at the third grade level, the relationships of OLMA/mathematics and reading/mathematics are stronger than OLMA/reading. At grades five and eleven the OLMA/reading relationship is strongest. The variance in achievement test scores which is accounted for by OLMA scores is shown under "r^2". For example, in grade three, variance in OLMA is said to account for 41% of the variation

in reading scores. So as expected, there is a strong relationship between ability scores and achievement scores for the group. The best that can be said about individual scores is found in the "Est." columns of Table 6. This indicates the accuracy with which one expects to predict achievement performance when mental ability performance is known. For example, at grade three, we say that when predicting reading, our estimate will be correct within $\pm$ 8.85 - a range of error of 17.70 points.

Cost of Administering Group Mental Ability Tests (OKC)

Cost Based on Fall 1979 Expenses

Research and Computer Services

Item	Rate	Hours	Total
Keypunch, employee	$4.75/hr	30	$ 142.50
10 Cases of Cards	$3.20/thous.	—	$ 320.00
Computer	$100/hr	25	$2500.00
Computer, operator	$6.35/hr	25	$ 168.75
Part time help, research	$3.13/hr	20	$ 62.60
Consumable tests	11.75/35pr.		$1307.50
		Total	$4501.35

School Personnel and Students

Students (two 50-minute periods x 17,500 students in grades 3, 5, 7, 9, 11)	28,334 hrs.
Teachers (based on an estimate of 580 teachers; 100 minutes for administering	97 hrs.
Estimated 5 minutes per test for scoring grade 3 tests)	283 hrs.
Counselors (receiving and distributing materials. In elementary schools would be principal or secretary)	1800 hrs.
Test Specialists (facilitating, distributing materials, analyzing and interpreting results)	200 hrs.
Total	30,714 hrs.

Cost if New Tests Are Purchased

Initial outlay for materials. The following figures suggest two options: 1) to test every grade, one through eleven; and, 2) to test every other grade, as we have been doing.

Option	Grades	Cost
1	1-11	$24,630.57 plus shipping
2	3, 5, 7, 9, 11	$10,078.57 plus shipping

Initial preparation, testing and scoring

Personnel and students

Test Specialist approx. 15 weeks	approx.	600 hrs.
Computer operator, approx. 4 weeks		160 hrs.
NCS operator		40 hrs.
Student (one class period x 17,500)		14,584 hrs.
Teachers		767 hrs.
Counselors		1,800 hrs.
	Total	17,951 hrs.

Maintenance costs with new tests

Materials grades 3,5,7,9,11		5,658.57
grades 1-11		$11,778.57
NCS operator		40 hrs.
Test Specialist		200 hrs.
Computer operator		160 hrs.
Students		14,584 hrs.
Teachers		767 hrs.
Counselors		
	Total	15,751 hrs.

Pending Legislation Affecting Testing In OKC

Section 4 of House Bill 1182 (known as Oklahoma School Testing Act of 1981) is quoted as follows:

"A. The State Board of Education shall:

1. Require a testing program in all school districts;

2. Designate the achievement, scholastic aptitude and physical performance tests to be used during the ensuing year;

3. Adopt regulations for the conduct and administration of testing program; and,

4. Make an annual report of the testing program and its results to the Legislature not later then January 15 of each year.

B. The State Board of Education may develop, publish and administer tests of its own devising, and the Board may utilize the expert services of any person or group of persons in public or private employment."

Input from Testing Committees

School Committees

During January of 1981, elementary, fifth year, middle and high school testing committees discussed uses of mental ability tests. They described the same uses given earlier in the "Purposes" section. Other points made by the committee included the following:

- Test scores which are more than two years old are sometimes used in order to obtain a score.
- Some teachers misunderstood the meaning of the term "mental ability" when responding to a district-wide survey.
- Some members suggested that mental ability testing be optional and that any teacher wanting to use the results would administer and score the test.
- If testing continues, students should be tested in fourth rather than fifth grade because it was believed that some students leave the district for fifth grade only.
- High school personnel expressed concern that students with problems are not identified before reaching ninth grade.

- A desire was expressed for more usable information than can be obtained from one overall score.
- The opinion that test results are misused was expressed.

Steering Committee

Prior to the February 18, 1981, meeting, members were asked to review a rough draft of the foregoing sections. At the meeting, members noted that pending legislation could overturn any decision regarding mental ability testing. It was further noted that, according to findings in this study, some of our uses of ability tests are actually misuses. Other ways of obtaining student information which study suggested were pointed out. These included: criterion-referenced tests, achievement tests, individual intelligence tests, and the "time to learn" concept. The following recommendations were made by the Testing Specialist:

- That the old group mental ability test (OLMA) not be used because norms are 23 years old.
- That the district not invest in a new group mental ability test this year. This would give time to learn the outcome of pending legislation and to explore other ways of gaining information about student ability.
- If and when a new test is purchased, the norming group should match that of the CAT. While it is understood that norming samples are made as representative as possible, the composition of every sample will be somewhat different.

- If and when a new test is purchased, it should be administered every year to provide an accurate assessment of current strengths and weaknesses.

These recommendations were summarized by the committee as follows:

1. The OKC district should stop administering the Otis-Lennon Mental Ability Test because the norms are 23 years old and because they are not validated with the CAT norms.

2. No new aptitude test should be purchased because of the high correlation between CAT and the companion aptitude test; i.e., the two tests measure the same thing.

3. The committee should look at aptitude testing again in the future after the legislature has made a decision on L.B. 1182.

4. Teachers should be given continuing assistance in using other sources of information.

References

Anastasi, Anne. Intelligence testing in the 1980's. Seminar transcript available from Scott, Foresman and Co., Glenview, Illinois, 1980.

Berger, B. A new interpretation of the I.Q. controversy. Public Interest, 1978, 50, 29-44.

Bloom, B.S. Human characteristics and school learning. New York: McGraw-Hill, 1976.

Bloom, Benjam J., et. al. Taxonomy of educational objectives handbook I: cognitive domain. New York: David McKay, Inc., 1956.

Blum, Jeffrey M. Pseudoscience and mental ability: the origins and fallacies of the I.Q. controversy. Monthly Review Press, New York, 1978.

Clarizo, Harvey F. In defense of the I.Q. test. School Psychology Digest, 1979, 8 (1) 82-87.

Coffman, W.E. Testing the disadvantaged: the test developer's point of view. AMEG Newsnotes, 1967, 3 (3).

Cottle, Thomas J. Going up, going down. In Paul L. Houts (ed.), The myth of measurability. New York: Hart Publishing Co., 1975.

Deutsch, M., Jishman, J.A., Kogan, L.S., North, A.D., Whitman, M. Guidelines for testing minority group children. Journal of Social Issues, 1964, 20, 129-145.

Dimengo, Carl. Basic testing programs used in major school systems throughout the United States in the school year 1977-78. Unpublished study, Akron Public Schools, Akron, Ohio: 1978.

Fine, Benjamin. The stronghold of the I.Q. Garden City, New York: Doubleday, 1975.

Flaugher, R. The many definitions of test bias. American Psychologist, 1978, 33, 671-679.

Gettinger, Maribeth, and White, Mary Alice. Which is the stronger correlate of school learning? Time to learn or measured intelligence? Educational Psychology, 1979, 71 (4), 405-411.

Glaser, Robert. Trends and research questions in psychological research on learning and schooling. Educational Researcher, 1979, Nov. 6-13.

Hall, Janie L. Results of a survey regarding uses of the Otis-Lennon Mental Ability Test. Unpublished study. Planning, Research, and Evaluation Department, Oklahoma City Public Schools, Oklahoma City, Oklahoma, 1980.

Hall, Janie L. Survey of large urban districts regarding use of mental ability tests. Unpublished study. Planning, Research, and Evaluation Department, Oklahoma City Public Schools; Oklahoma City, Oklahoma, 1981.

Hall, Janie L. and Dutcher, Louise. District profile. Oklahoma City, Oklahoma. Oklahoma City Public Schools, 1980.

Horn, John L. Trends in the measurement of intelligence. Intelligence, 1979, 3, 229-240.

Jackson, Frances Mae. Group standardized testing results. Oklahoma City, Oklahoma. Oklahoma City Public Schools. 1973, 1974, 1975, 1976, 1977, 1978, 1979.

Jacobs, James N. Testing for the urban center in the '80s: three challenges. Seminar transcript available through Scott, Foresman and Co., Glenview, Illinois, 1980.

Jacobs, P.I. and Vandeventer, M. The learning and transfer of double classification skills: a replication and extension. Journal of Experimental Child Psychology, 1971, 12, 140-157.

Jensen, Arthur R. Bias in mental testing. New York: Free Press, 1980.

Jones, W. Paul, and De Blassie, Richard R. Social class contrasts in short term predictability of grade achievement. Journal of Educational Research, 1971, 65 (1).

Kovacs, Frnak W, director. Survey of school programs and procedures 1980. National Education Association, 1201 Sixteenth St, N.W., Washington D.C. 20036, September 1980.

Lennon, Roger T. Perspective on intelligence testing. NCME Measurement in Education, 1978, 9 (2) 6-8.

Lloyd, R. Grann. Effective and ineffective uses of tests as indicators of student achievement. High School Journal, 1979, Nov., 76-79.

Mercer, J. Technical manual: SOMPA. New York: Psychological Corp, 1979.

Neisser, U. The concept of intelligence. Intelligence, 1979, 3, 217-227.

Olson, D.R. Culture, technology, and intellect. In L.B. REsnick (ed.). The nature of intelligence. Hillsdale, N.J.: Erlbraum, 1976, ch 10.

Olton, R.M. and Crutchfield, R.S. Developing the skills of productive thinking. In P.H- Mussen, J. Langer, and M. Covington (eds.). Trends and issues in developmental psychology. New York: Holt, Rinehart and Winston, 1969.

Otis, Arthur S. and Lennon, Roger T. Technical handbook, Otis-Lennon Mental Ability Test. New York. Harcourt, Brace and World, Inc. 1969

Otis, Arthur S. and Lennon, Roger T. Otis-Lennon School Ability Test, Manual for Administering and Interpreting. New York. Harcourt, Brace, and Jovanovich, Inc. 1979.

Reschly, Dan. Progress report of school psychology and current directions in the field. Oklahoma School Psychological Association Newsletter, Jan. 1981, Vol 5, (1).

Resnick, B. The future of I.Q. testing in education. Intelligence, 1979 3, 241-253.

Resnick, L.B. and Glaser R. Problem solving and intelligence. In L.B. Resnick (ed.). The nature of intelligence. Hillsdale, N.J.: Erlbraum, 1976, ch 11.

Robertson, Gary T. and Mitchell, Blythe C. Stability of Mental Ability Test Performance over three - and six-year intervals. Unpublished paper, The Psychologican Corp., 1975.

Spearman, Charles. General intelligence objectively determined and measured. American Journal of Psychology, 1904, 15, 201-293.

Steffensen, M.S., Jogdes, Cl. and Anderson, R.D. Across-cultural perspective on reading comprehension. Technical report No. 97, Center for the study of reading. Champaign, Illinois, 1978

Sternberg, Robert J. Six authors in search of a character: a play about intelligence tests in the year 2000. Intelligence, 1979, 3, 283-293.

Thorndike, R.L. (Editor). Educational measurement (2nd. ed.). Washington, D.C.: National Council on Education, 1971.

Turnbull, William W. Intelligence testing in the year 2000. Intelligence, 1979, 3, 275-282.

Vandiver, Phillip L., and Vandiver, Stella Sue. A "nonbiased assessment" of intelligence testing. Educational Forum, 1979, Nov., 97-108.

Wallbrown, Fred H. Intelligence testing. The Directive Teacher, 1979, Fall, p 20.

Weber, George. Uses and abuses of standardized testing in the schools. (Occasional Papers, No. 22, Council for Basic Education). Washington, D.C.: 725 Fifteenth Street N.W., 1978

Wick, John W., Smith, Jeffrey K., Breggs, Donald L., Cawley, John F., Mouw, John T., and Cawley, Louise J. Developing Cognitive Abilities Test, Teacher's Manual. Glenview, Illinois: Scott, Foresman and Co., 1980.